



Hypothesis report

Project 1: XML Schema Definition

Authors:

Siéwé Kouéta ANICET : 00364245

Professor:

Stijn VANSUMMEREN



UNIVERSITÉ LIBRE DE BRUXELLES

ULB

April 21, 2019

Contents

1	Introduction	2
2	Structure of XML documents	2
2.1	<ScientificProducts> tag	2
2.1.1	<book> tag	2
2.1.2	<journals> tag	2
2.2	The <LeisureProducts> Tag	3
2.2.1	the <book> tag	3
2.2.2	the <periodicals> tag	3
3	Constraint of the xml schema.	3
3.1	The type of fields	4
3.2	Restrictions	4
4	difference between bookshop.xml and bookshop2.xml	5
5	Annex	5
5.1	<authors> tag	5
5.2	<Editor> tag	5
5.3	<Article> tag	6
5.4	XML tree	6
	References	10

1 Introduction

This report aims to give an explanation detailing the choice of our hypotheses and the constraints to use to build the XML shema document "**bookshop_xml_shema.xsd**" as well as the tree of the XML documents to associate "**bookshop.xml**" and "**bookshop2.xml**".

2 Structure of XML documents

Nous avons utiliser la version 1.0 des document xml coder en UTF-8,

The root element of XML documents is the **<bookshop>** tag. We assumed that the departments are split into two separate tags. in the root element, each department is represented by a **<ScientificProducts>** and **<LeisureProducts>** tag.

2.1 **<ScientificProducts>** tag

this tag contains two sub-tags: **<book>** and **<journals>**

2.1.1 **<book>** tag

It contains two attributes:

- a **'type'** attribute that specifies the type of free (mandatory attribute)
 - a **'ISBN'** attribute which gives a reference to this book (optional attribute)
- This tag encapsulates 6 other tags, so 4 are required, structure in the following order.

- **<title>**
- **<ListingAuthors>** / **<ListingEditors>** (in turn encapsulates a **<authors>** / **<editors>** tag: described in detail in the appendix)
- **<publisher>**
- **<YearPublication>**
- **<abstract>** (optional)
- **<edition>** (optional)

2.1.2 **<journals>** tag

it has no attributes and encapsulates 8 other tags, so 6 are mandatory, and 2 optional struct in the following order:

- **<title>**
- **<volume>**
- **<number>**
- **<NumRedactor>** / **<ListingAuthors>** (in turn encapsulates a **<authors>** tag: described in detail in the appendix)

- <YearPublication>
- <publisher> (optional)
- <ImpactFactor > : optional, and contains a attribute 'Year' attribute
- <contents>: encapsulates an <article> tag, described in detail in the Annex section)
-

2.2 The <LeisureProducts> Tag

It contains the sub-tags: <book> and <periodicals>,

2.2.1 the <book> tag

It contains two attributes:

- a 'type' attribute that specifies the type of free (mandatory)
- a 'genre' attribute that gives the genre of the book. (mandatory)

This tag encapsulates 6 other tags, so 4 are required to structure in the following order :

- <title>
- <ListingAuthors>: encapsulates an <authors> tag, which is described in detail in Annex section
- <publisher>
- <YearPublication>
- <edition> (optional)
- <PageNum> (optional)

2.2.2 the <periodicals> tag

It encapsulates 3 other tags:

- <title>
- <ListingAuthors>: encapsulates an <authors> tag, which is described in detail in Annex section
- <title>
- <price currency=""> : who has a mandatory attribute 'currency'

3 Constraint of the xml schema.

To structure our xml documents, they must obey a set of rules and constraints defined in xml shema.

3.1 The type of fields

type	positiveInteger	float	string	gYear
comment	Contains only positive number negative(numbers are not allowed)	Contains integer or decimal	Contains string of characters	Contains a date like year
element	<StartPage> <EndPage> <NumRedactor> <number> <numero> <PostalCode> <volume>	<price>	<title> <name> <FirstName> <rue> <city> <country> <Phone> <email> <publisher> <abstract> <edition>	<YearPublication>

3.2 Restrictions

- **<book type=" " isbn=" ">** : of the scientific department (**<ScientificProducts>** element). To allow the type of free couplet, there are two attributes: 'type' and 'isbn'. the 'type' attribute is mandatory and can take only two possible values {'scientific' or 'leisure'}. The 'isbn' attribute is optional and is of type ID, so must have a unique value in the document
- **<book type=" " genre=" ">**: of the leisure department (**<LeisureProducts>** element). To distinguish the type of free, there are two attributes: 'type' and 'genre'. the 'type' attribute is mandatory and can take only two possible values {'scientific' or 'leisure' }. The 'genre' attribute is as obligatory and can take as 4 possible values: {'thriller', 'horror', 'sci / fi', 'romance' or 'literature' }
- **<PostalCode>**: the postal code must obligatorily contain 4 digit included between [0-9]
- **<Phone type=" " locality=" ">** : each phone element has two attributes 'type' and 'locality'. The 'type' attribute can only take three possible values {'HousePhone', 'cellphone' or 'OfficePhone' }. The attribute 'locality' is optional and can only take 4 possible values {'BEL', 'FRA', 'CMR', 'USA' }

Note: The possible values of the 'locality' attribute are just one example. It is of course possible to extend this list of possibilities in the `CountryCode.xsd` file.

- **<email type=" ">** : each email element has a 'type' attribute. The 'type' attribute can only take three possible values {'personal', or 'professional' }
- **<authors ID=" ">** : we assumed that each author should be uniquely identified. For this we add an ID attribute in this tag. This attribute is mandatory and is of type 'ID', so must have a unique value in the document. In addition, this tag must appear at least once and as many times as possible in the **<ListingAuthors>** tag.
- **<person sexe=" ">** : we have assumed that there are two types of person (Male and Female). in order to make this distinction, we have added the attribute 'sex' which is mandatory and can only take two possible values: {'M' or 'F' }

4 difference between bookshop.xml and bookshop2.xml

To be able to provide two different files that will respect the rules of our xml schema, we created these two documents `bookshop.xml` and `bookshop2.xml`.

In the `bookshop.xml` document, we have included all the nodes, even the optional nodes. But in `bookshop2.xml` we have removed all the optional nodes: so nodes like `<edition>`, `<PageNum>`, `<abstract>`, `<ImpactFactor>` do not appear in `bookshop2.xml`.

In our xml schema, wherever there was the choice of appearance between two elements, were replaced by the second choice in the `bookshop2.xml` document. For example in the `bookshop.xml` node, `<ListingAuthors>` is used. But in `bookshop2.xml`, it is replaced by the `<ListingEditors>` node.

5 Annex

5.1 `<authors>` tag

This tag encapsulates another `<person>` tag, because we wanted to distinguish the genre of the author (male or female). The tag `<person sex = "">` therefore has a `'sex'` attribute. `<person>` encapsulates 5 other tags all mandatory:

- `<name>`: name of the author
- `<FirstName>`
- `<adresse>` : to define how to efficiently handle an address, this tag encapsulates another 5 tags:
 - `<numero>` : building number
 - `<rue>` : name of the street or avenue
 - `<PostalCode>` : ZIP code
 - `<city>`
 - `<country>`
- `<Phones>`: it encapsulates a single tag
 - `<Phone type="" locality="" >` : which has two attributes `'type'` (required) and `'locality'` (optional). This makes it possible to accurately distinguish the author's personal or professional telephone number, as well as the country of the number.
- `<emails>` : it encapsulates a single tag
 - `<email type="" >` : which has a `'type'` attribute (optional). This allows you to accurately distinguish personal email or professional author

5.2 `<Editor>` tag

It encapsulated another 4 tag structure in the following order:

- `<name>` : editor's name

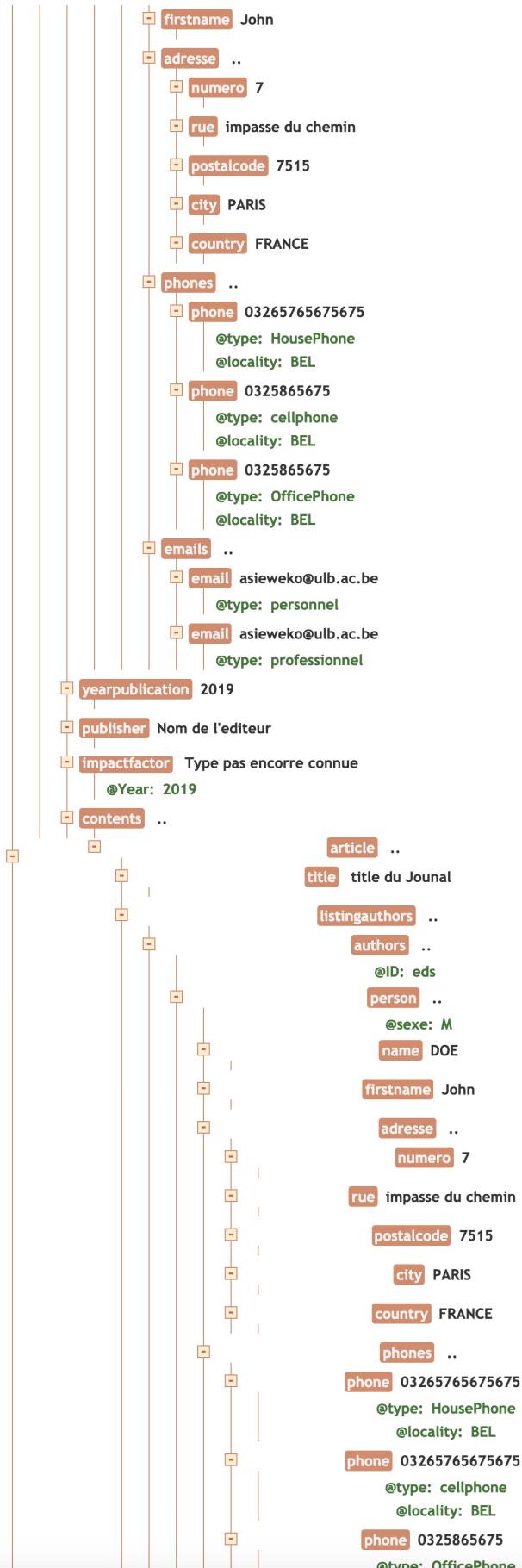
- <adresse>: which encapsulates in turn 5 other tags:
 - <numero> : building number
 - <rue> : name of the street or avenue
 - <PostalCode> : ZIP code
 - <city>
 - <country>
- <Phones>: it encapsulates a single tag
 - <Phone type="" locality="" > : which has two attributes 'type' (required) and 'locality' (optional).
- <emails> : it encapsulates a single tag
 - <email type="" > : which has a 'type' attribute (optional).

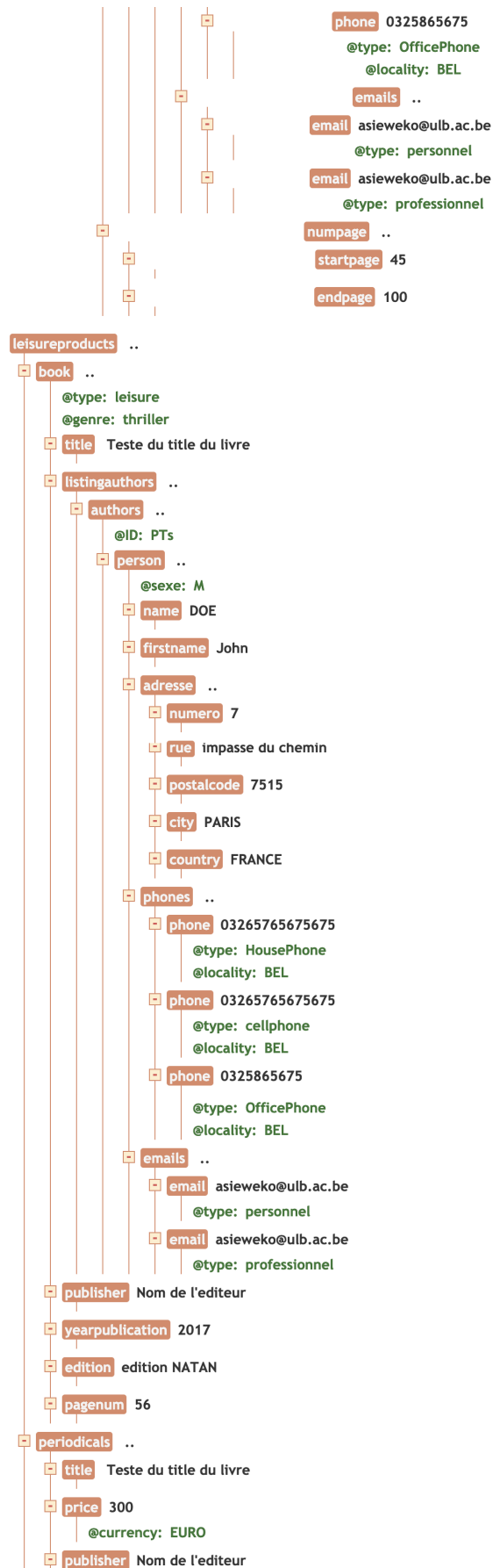
5.3 <Article> tag

- <title> :title of the article
- <ListingAuthors> :which in turn encapsulates the <authors> tag (described in section 1 of the Annex)
- <NumArticle> / <NumPage> :(The <NumPage> tag encapsulates two other <Start-Page> and <EndPage> tags)

5.4 XML tree







References